

SEQUENCE LISTING

5 (1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Zeneca MOGEN
(B) STREET: Einsteinweg 97
(C) CITY: Leiden
(E) COUNTRY: The Netherlands
(F) POSTAL CODE (ZIP): 2333 CB
(G) TELEPHONE: (31) 71-5258282
(H) TELEFAX: (31) 71-5221471

(ii) TITLE OF INVENTION: New constitutive plant promoters

(iii) NUMBER OF SEQUENCES: 22

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: EP 97203912.7
(B) FILING DATE: 12-DEC-1997

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 520 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GA CTGAAGTG TGAAGGTGGA GATTATGTAT TCACTTGTTC ATTTGGTATA CATTCTATGT 60
50 AAGGTTCAAT TATTTACGTT ATATAATTAT AATGGAGTAA TTTACAGTAA TTGGGTAAAA 120
ATGGTTTGAT TCGGTCAGGT TGATACGGTT TGGAAGTTAA ACCCGGCCTA GATATGATGT 180
TACAACCAAGT CCACATCTTT TATGATTTTA GTGGAACAAA CGAAGAGTTA TTTAGACGAT 240
55 ACAACAAGG TCCGAATAAG TGTGAGCTGT CCCAAGTAAG ACCACGTAAT ACTCACCTCA 300
ACAAGATAGT GTTCTTAAAG TGTGTCAAAC ACAATCACAC ACACACAAAT CATAAAACAC 360
60 AAAGACGATA ATCCATCGAT CCACAGAATA GACGCCACGT GGTAGATAGG ATTCTCACTA 420
AAAAGTTCTC ACCTTTTAAT CTTTCTCCAC GCCATTTCCA CAAGCCATAA TCCTCAAAAA 480
TCTCAACTTT ATCTCCCAA ACACAAATCT AGAAACCATG 520

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 300 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CCCACTACAA TGAATTTGTT CGTGAACATAT TAGTTGCGGG CTTGGCATC CGACTACCTC 60
TGGCGCAATA TTATATTCCT TGGGCCACCC GTGAACCCAA TTTCGCCTAT TTATTCATTA 120
CCCCCATTA CATTGAAGTA GTCATGATGG GCCTGCAGCA CGTTGGTGAG GCTGGCACAA 180
CTCATCCATA TACTTTCTGA CCGGATCGGC ACATTATTGT AGAAAACGCG GACCCACAGC 240
GCACTTTCCA AAGCGGTGCC GCGTCAGAAAT GCGCTGGCAG AAAAAAATTA ATCCAAAAGT 300

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 840 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGATCCGAGC TTGCATGCCC CCACTACAAT GAATTTGTTC GTGAACATAT AGTTGCGGGC 60
CTTGGCATCC GACTACCTCT GCGGCAATAT TATATTCCTT GGGCCACCCG TGAACCCAAT 120
TTGCCTATT TATTCATTAC CCCCATTAAC ATTGAAGTAG TCATGATGGG CCTGCAGCAC 180
GTTGGTGAGG CTGGCACAAAC TCATCCATAT ACTTTCTGAC CGGATCGGCA CATTATTGTA 240
GAAAACGCGG ACCCACAGCG CACTTTCCAA AGCGGTGCCG CGTCAGAAAT CGCTGGCAGA 300
AAAAAATTAA TCCAAAAGTG ACTGAAGTGT GAAGGTGGAG ATTATGTATT CACTTGTTGA 360
TTTGGTATAC ATTCTATGTA AGGTTCAATT ATTTACGTTA TATAATTATA ATGGAGTAAT 420
TTACAGTAAT TGGGTAAAAA TGGTTTGATT CGGTCAGGTT GATACGGTTT GGAAGTTAAA 480
CCCGGCCTAG ATATGATGTT ACAACCAAGTC CACATCTTTT ATGATTTTAG TGGAACAAAC 540

GAAGAGTTAT TTAGACGATA CAAACAAGGT CCGAATAAGT GTGAGCTGTC CCAAGTAAGA 600
CCACGTAATA CTCACCTCAA CAAGATAGTG TTCTTAAAGT GTGTCAAACA CAATCACACA 660
5 CACACAAATC ATAAAACACA AAGACGATAA TCCATCGATC CACAGAATAG ACGCCACGTG 720
GTAGATAGGA TTCTCACTAA AAAGTTCTCA CCTTTTAATC TTTCTCCACG CCATTTCCAC 780
AAGCCATAAT CCTCAAAAAT CTCAACTTTA TCTCCCAAAA CACAAATCTA GAAACCATGG 840

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 477 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Arabidopsis thaliana*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGATCCTGCA GCGATTTTCAT TTTAGATTCT CAAAAATATT CTCAGATGTG TGGGATTTGA 60
GTAGAGTTTA TGTTCGCTTG GCATGATTG AATAGTATGC AAGATTTTGT AGATTTTGCA 120
35 TTCGTTTCATG TGTGTATGTG TGATTGTAGC TTGATATGAT TTAACCTGTT AGTTAAATGT 180
GCATAGACAA TAAGTAACAT ACGAAGCGAG TCACTAAGCA TAAGAGTCAA CTTGTTTTGC 240
40 TGAAAAGATA TCACTTATGA TTTTCGAATC ATTTTAGCTT TTTTGTCAC TGAAGCTTAAT 300
GATTCTTCTG AAATTCGATT CTTTGTTTGG TTTATGTCAC ATTCTTTAGA ATTGAGAATC 360
TAAGAAATGC TTACAGGATA TGGTGAAACT ATTCTTTTAA GATAGCATGA TGCTTCTTTT 420
45 ATGATTCTAC AGTGGCTAAG TCATTTTTTT TTTGTTCTAT TCTTTGTAGC ACCATGG 477

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

AGATTTGTAT TGCAGCGATT TCATTTTAG

29

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

ATCTGGTCAC AGAGCTTGTC

20

(2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GTCTCCATGG TGCTACAAAG AATAG

25

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CGGGATCCTG CAGCGATTTC ATTTAG

27

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
ACATGAACGA ATGCAAAATC TC 22

(2) INFORMATION FOR SEQ ID NO: 10:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
AGATTTTGCA TTCGTTTCATG TG 22

(2) INFORMATION FOR SEQ ID NO: 11:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
TGTAAGCATT TCTTAGATTC TC 22

(2) INFORMATION FOR SEQ ID NO: 12:

40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:
AAGAAATGCT TACAGGATAT GG 22

(2) INFORMATION FOR SEQ ID NO: 13:

55 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs

60

65

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GACAAGCTTG ATCCCATGGT GCTACAAAGA ATAG

34

15 (2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: cDNA

25 (iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

30

AGTGGTACCA TCATAATACT CATCCTCCTT C

31

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

35

40

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CGAAGCTTTA CAAATCTAAT TTCATCACTA AATCGGA

37

50

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: cDNA

60

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

65 GCTGCAATAC AAATCTAATT TCATCACTAA ATCGG

35

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 876 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Arabidopsis thaliana

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATCATAATAC TCATCTCCT TCTCAAGGTT CGTACGTATT ATCAATATCT AGTATATACT 60
 TGTCTTTGTT CTATGCTTTA TATCATCATT TTATGACAAA AAATGATTAA GGTCTTAGTT 120
 AATGATTATG TATATGTGAA ACTTATATTT AGGGGCACAA TTTAATTTTCG TATGATAAAT 180
 GTCTAGTTAG CTTTATGTAC TTATCATAAA AACCTTAGTG TTTATCGCAA TACTTTTCAA 240
 ATATAGTGTA GAATCATAAT GGTCCCACTG TCATTATGTT TGATGCAAAT CTATTTGGAT 300
 TTTGTTGGAT AATAAACCGA TGACGTGGAC CAGACCAGTA GCTATAAGAT TTGGTTCACA 360
 TAGAAATTTT TTATAAGATA ATGTATCTAG GTTTGCTTAT GATTATACAT GTGATATTTA 420
 ATACATGGCA CAGGTTTCGTC GAGTTTCACA GCCATAGGTA CAATAGAAGG CAAATTCGAT 480
 TGTGGTTATC TGGTAAAAGT TAAGTTGGGC TCAGAGATTC TTAACGGCGT TCTTTATCAT 540
 TCGGCCCAGC CCGGCCCATC ATCATCTCCA ACCGCTGTTC TAAACAATGC CGTTGTACCT 600
 TATGTTGAAA CTGGGAGGAG ACGGCGTCGT TTAGGTAAAA GACGAAGAAG CAGACGCAGA 660
 GAAGATCCGA ATTACCCGAA ACCGAACCGG AGCGGTTACA ATTTCTTCTT TGCTGAGAAA 720
 CATGCAAGC TCAAATCACT TTATCCCAAC AAGGAGAGAG AGTTTACGAA ACTTATCGGA 780
 GAATCGTGGA GCAATCTCTC TACCGAAGAA CGAATGGTAA CAAATTATCT TTAAACCGT 840
 TACCGATTTA GTGATGAAAT TAGATTTGTA GTAAAT 876

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1357 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GGATCCCCGG GTACCATCAT AATACTCATC CTCCTTCTCA AGGTTCTGAC GTATTATCAA 60
 5 TATCTAGTAT ATAATTGTCT TTGTTCTATG CTTTATATCA TCATTTTATG ACAAAAAATG 120
 ATTAAGGTCT TAGTTAATGA TTATGTATAT GTGAACTTA TATTAGGGG CACAGTTTAA 180
 TTTCTGATGA TAATTGTCTA GTTAGCTTTA TGTACTTATC ATAAAAACCT TAGTGTTTAT 240
 10 CGCAATACTT TTCAAATATA GTGTAGAATC ATAATGGTCC CACTGTCATT ATGTTTGATG 300
 CAAATCTATT TGGATTTTGT TGGATAATAA ACCGATGACG TGGACCAGAC CAGTAGCTAT 360
 15 AAGATTTGGT TCACATAGAA ATTTTTTATA AGATAATGTA TCTAGGTTTG CTTATGATTA 420
 TACATGTGAT ATTTAATACA TGGCACAGGT TCGTCGAGTT TCACAGCCAT AGGTACAATA 480
 GAAGGCAAAT TCGATTGTGG TTATCTGGTA AAAGTTAAGT TGGGCTCAGA GATTCTTAAC 540
 20 GCGGTTCTTT ATCATTTCGC CCAGCCCGGC CCATCATCAT CTCCAACCGC TGTTCTAAAC 600
 AATGCCGTTG TACCTTATGT TGAACTGGG AGGAGACGGC GTCGTTTAGG TAAAAGACGA 660
 25 AGAAGCAGAC GCAGAGAAGA TCCGAATTAC CCGAAACCGA ACCGGAGCGG TTACAATTTT 720
 TTCTTTGCTG AGAAACATTG CAAGCTCAAA TCACTTTATC CCAACAAGGA GAGAGAGTTT 780
 ACGAACTTA TCGGAGAATC GTGGAGCAAT CTCTCTACCG AAGAACGAAT GGTAACAAAT 840
 30 TATCTTTTAA ACCGTTACCG ATTTAGTGAT GAAATTAGAT TTGTATTGCA GCGATTTTAT 900
 TTTAGATTCT CAAAAATATT CTCAGATGTG TGGGATTGTA GTAGAGTTTA TGTGCGTTG 960
 35 GCATGATTG AATAGTATGC AAGATTTTGT AGATTTTGCA TTCGTTTATG TGTGTATGTG 1020
 TGATTGTAGC TTGATATGAT TTAACCTGTT AGTTAAATGT GCATAGACAA TAAGTAACAT 1080
 ACGAAGCGAG TCACTAAGCA TAAGAGTCAA CTGTTTTGTC TGAAAAGATA TCACTTATGA 1140
 40 TTTTCGAATC ATTTAGCTT TTTGTCACT TGAGCTTAAT GATTCTTCTG AAATTCGATT 1200
 CTTTGTGTTGG TTTATGTCAC ATTCTTTAGA ATTGAGAATC TAAGAAATGC TTACAGGATA 1260
 45 TGGTGAACT ATTCTTTTAA GATAGCATGA TGCTTCTTTT ATGATTCTAC AGTGGCTAAG 1320
 TCATTTTTTT TTTGTTCTAT TCTTGTAGC ACCATGG 1357

(2) INFORMATION FOR SEQ ID NO: 19:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

65

GTCTTGATCA AATCTAATTT CATCACTAAA TCGG

34

(2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TTAGATTTGT ACAAGACCCT TCCTCTATAT AAGG

34

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

TTCCCAGTCA CGACGTTGT

19

(2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1006 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GGATCCCCGG GTACCATCAT AATACTCATC CTCCTTCTCA AGGTTCGTAC GTATTATCAA

60

TATCTAGTAT ATACTTGTCT TTGTTCTATG CTTTATATCA TCATTTTATG ACAAAAAATG

120

ATTAAGGTCT TAGTTAATGA TTATGTATAT GTGAACTTA TATTTAGGGG CACAGTTTAA

180

TTTCGTATGA TAATTGTCTA GTTAGCTTTA TGTACTTATC ATAAAAACCT TAGTGTTTAT

240

CGCAATACTT TTCAAATATA GTGTAGAATC ATAATGGTCC CACTGTCATT ATGTTTGATG

300

CAAATCTATT TGGATTTTGT TGGATAATAA ACCGATGACG TGGACCAGAC CAGTAGCTAT

360

AAGATTGGT TCACATAGAA ATTTTATA AGATAATGTA TCTAGGTTTG CTTATGATTA 420
TACATGTGAT ATTTAATA TGGCACAGGT TCGTCGAGTT TCACAGCCAT AGGTACAATA 480
5 GAAGGCAAAT TCGATTCTGG TTATCTGGTA AAAGTTAAGT TGGGCTCAGA GATTCTTAAC 540
GGCGTTCTTT ATCATTCTGG CCAGCCCGGC CCATCATCAT CTCCAACCGC TGTCTAAAC 600
10 AATGCCGTTG TACCTTATGT TGAACTGGG AGGAGACGGC GTCGTTTAGG TAAAAGACGA 660
AGAAGCAGAC GCAGAGAAGA TCCGAATTAC CCGAAACCGA ACCGGAGCGG TTACAATTTC 720
TTCTTTGCTG AGAAACATTG CAAGCTCAAA TCACCTTATC CCAACAAGGA GAGAGAGTTT 780
15 ACGAACTTA TCGGAGAATC GTGGAGCAAT CTCTCTACCG AAGAACGAAT GGTAACAAAT 840
TATCTTTTAA ACCGTTACCG ATTTAGTGAT GAAATTAGAT TTGTACAAGA CCCTTCCTCT 900
20 ATATAAGGAA GTTCATTTC TTTGGAGAGG ACACGTATTT TTACAACAAT TACCAACAAC 960
AACAAACAAC AAACAACATT ACAATTACTA TTTACAATTA CCATGG 1006